

Title (en)
Construction of a large co-occurrence data file

Title (de)
Erstellung einer großen Datei von Kookkurrenzdaten

Title (fr)
Construction d'un fichier de données de cooccurrence de grande taille

Publication
EP 1939767 A1 20080702 (FR)

Application
EP 07122997 A 20071212

Priority
FR 0655868 A 20061222

Abstract (en)
The size of co-occurrence data file is divided into blocks occupying memory space at most equal to size of a buffer block of a first memory (MC). Each block of the file is processed by reading corpus (C) and incrementing by one unit a frequency count associated with objects of each block if those objects are grouped in the read corpus to satisfy a co-occurrence criterion. The co-occurrence data and the associated non-null frequency counts corresponding to said co-occurrence data are transferred from the buffer block of first memory to co-occurrence data file in second memory (DSQ). Independent claims are also included for the following: (1) a data processing system; (2) a computer arrangement performed in a data processing system; and (3) a method for automatic reformulation of a search request in a search application.

Abstract (fr)
Un système informatique pour construire un fichier de données de cooccurrence comporte des première et deuxième mémoires. Un module (DI) dimensionne la taille du fichier par un recensement d'objets distincts dans un corpus (C). Un module (DV) divise la taille du fichier en blocs (B ij) occupant un espace mémoire sensiblement au plus égal à la taille de la première mémoire, chaque bloc faisant correspondre des objets entre eux. Un module (TR) traite chaque bloc du fichier en lisant le corpus et en incrémentant d'une unité un compte associé à des objets du bloc si ces objets sont groupés dans le corpus en satisfaisant à un critère, chaque groupement d'objets correspondant à une donnée de cooccurrence. A la fin de la lecture du corpus, le bloc traité dans la première mémoire est transféré vers le fichier dans la deuxième mémoire.

IPC 8 full level
G06F 17/30 (2006.01)

CPC (source: EP US)
G06F 16/31 (2018.12 - EP US); **G06F 40/284** (2020.01 - EP US)

Citation (search report)
• [YA] US 5675819 A 19971007 - SCHUETZE HINRICH [US]
• [YA] US 6314419 B1 20011106 - FAISAL MOHAMMAD [US]
• [A] US 6411962 B1 20020625 - KUPIEC JULIAN M [US]
• [XY] SHIEH W-Y ET AL: "Inverted file compression through document identifier reassignment", INFORMATION PROCESSING & MANAGEMENT, ELSEVIER, BARKING, GB, vol. 39, no. 1, January 2003 (2003-01-01), pages 117 - 131, XP004397215, ISSN: 0306-4573
• [XY] MCKELLAR A C ET AL: "Organizing Matrices and Matrix Operations for Paged Memory Systems", COMMUNICATIONS OF THE ASSOCIATION FOR COMPUTING MACHINERY, ACM, NEW YORK, NY, US, vol. 12, no. 3, March 1969 (1969-03-01), pages 153 - 165, XP007902169, ISSN: 0001-0782

Cited by
EP2256655A1; FR2922337A1; WO2009053643A3; US8453249B2; US9785635B2

Designated contracting state (EPC)
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC MT NL PL PT RO SE SI SK TR

Designated extension state (EPC)
AL BA HR MK RS

DOCDB simple family (publication)
EP 1939767 A1 20080702; US 2008154992 A1 20080626

DOCDB simple family (application)
EP 07122997 A 20071212; US 95549307 A 20071213